Proteins at atomic resolution

Zbigniew Dauter, Victor S Lamzin and Keith S Wilson

European Molecular Biology Laboratory, Hamburg, Germany and University of York, York, UK

Experimental advances in data collection, including bright sources, cryogenic cooling and two-dimensional detectors, have made it tractable to record data to beyond 1.2 Å for several proteins, yielding high-accuracy models and fine details of structure. For small metalloproteins, atomic-resolution data have enabled *ab initio* solution of the phase problem.

Current Opinion in Structural Biology 1995, 5:784–790

Introduction

This year marks the 100th anniversary of the discovery of X-rays by Röntgen. Since that discovery the application of X-rays to a broad spectrum of problems, including medical imaging and molecular structure determination, has become standard. The latter field includes a variety of techniques, such as small angle scattering, X-ray spectroscopy, powder diffraction and single crystal X-ray diffraction, with which we are here concerned.

Why are X-rays so well suited to the determination of atomic structures in three dimensions? The answer lies in a combination of their physical attributes. Firstly, X-rays can be generated from appropriate metal anodes with wavelengths of ~1.0 Å, ideal for resolving the positions of individual atoms. Secondly, X-rays penetrate deeply into crystals without being totally absorbed, unlike electrons. Thirdly, they have a reasonable cross-section of interaction with the electrons around the atomic centres (the diffraction efficiency for organic structures is approximately 1 in 10^{-4}) much better than do neutrons. Finally, they are cheap and easy to produce in a conventional laboratory, although intense sources of synchrotron radiation again enter the 'big science' level of funding.

The great advantage of X-ray crystallography is that for normal small molecules data can be recorded to truly atomic resolution, beyond 1.0 Å, allowing the positions of the ordered atoms in the structure to be accurately identified with a typical coordinate error of ~0.002 Å. Moreover, the accuracy relates as much to long as to short interatomic distances within the molecule, unlike many spectroscopic methods which give only short-distance information. Given these attributes, single crystal X-ray techniques have become routine in the small-molecule field. Tens of thousands of structures have been determined, both in academia and in industry. The major limitation to the scope of X-ray studies is the phase problem, as only the amplitudes, and not the phases, of the diffracted rays can be recorded experimentally. This problem is essentially solved for small structures.

Particular problems attend the solution of larger structures, involving not only their sheer size but especially their degree of disorder, factors which make the collection of atomic-resolution data a challenging prospect. The lack of such data causes difficulties at all stages of a structure analysis, from data collection through structure solution to refinement. Several advances during recent years, including two-dimensional (2D) detectors, high-intensity sources and cryogenic cooling, have meant that atomic resolution can be recorded for at least a small subset of well ordered protein crystals.

Although we feel that a review of progress in this field is timely, it is not possible to refer to an extensive set of published results. With the exception of a small number of pioneering studies, such as those involving avian pancreatic polypeptide [1], crambin [2], rubredoxin [3] and insulin [4], data on other structures have been collected within the past three years or so and most are in the process of detailed refinement. We cannot therefore provide a true survey of the work of all groups in the field. Instead, we will give a largely subjective description of our own studies at the European Molecular Biology Laboratory (EMBL) in Hamburg, including the projects of a number of groups that visit the facility. We hope that such an offering will stimulate others to make best use of their crystals and to attain atomic resolution wherever possible.

Advantages of atomic resolution

We heuristically define atomic resolution according to George Sheldrick [5]: the data should extend to at least 1.2 Å, and at least 50% of data in the outer resolution shell should have intensities of $>2\sigma$, which roughly corresponds to a merging R factor of ~25%. Data at the resolution edge with errors above this level may prove to be

Abbreviations

AKP—Automatic Refinement Procedure; **CCD**—charge-coupled device; **2D**—two-dimensional; **EMBL**—European Molecular Biology Laboratory; **MSD**—Macromolecular Structural Database; **NCS**—non-crystallographic symmetry.

important in many applications, for example, where data are sparse, or in solving the phase problem using direct methods.

Atomic-resolution data confer two important possibilities. Firstly, for small structures, the phase problem can be solved directly using either Patterson or direct methods. This has also been shown to hold true for small metalloproteins but not for proteins that lack heavy-atom centres. Secondly, they allow comprehensive least-squares refinement of the structure with anisotropic atomic temperature factors: at 1.0 Å resolution, for example, there are still more than five observations per parameter. The positions of most of the H atoms can be identified in the electron density. The final R factor can be as low as 2% and the residual results largely from the neglect of bonding and lone-pair electrons in the spherical-atom approximation.

Problems of data collection for large structures

Data collection for macromolecules presents two problems, inherent in the nature of the crystals and responsible for all subsequent difficulties in the crystallographic analysis. The first is that the number of X-ray data increases with the cube of the average cell dimension while the average intensity of the reflections decreases (See Table 1). Thus, data collection to atomic resolution is a daunting task even for small proteins. This problem affects all data, both at high and at low resolution. Assuming that the crystals are well ordered, the problem boils down to improving the signal-to-noise ratio, which requires improved counting statistics. This can be achieved by using a higher incident intensity on the crystal combined with simultaneous recording of all diffracted reflections.

Table 1. The problem of data collection to atomic resolution (~1 Å) for structures of increasing size*.

Structure	Cell edge	No. of reflections	Mean intensity
Small organic	10	1000	1
Supramolecule	20	8000	1/8
Protein	50	125 000	1/125 000
Virus	300	27 000 000	1/27 000 000

*Most proteins and all viruses do not in practice diffract to 1 Å resolution. The difficulties arise from the number of reflections increasing as the cube of the average cell dimension and their average intensity decreasing also as the cube.

The higher intensity can be achieved by using modern rotating-anode generators; however, the use of synchrotron radiation is more effective. Rotating anodes may provide up to 10^9–10^{10} photons mm^{-2} s^{-1}, third-generation synchrotrons approx. 10^{12}, and fourth-generation sources such as the European Synchrotron Radiation Facility or the Advanced Photon Source should provide at least a further two orders of magnitude. In contrast, efficient recording of the data poses a challenge. Two-dimensional detectors have become standard in most laboratories, and the current detector of choice for the home laboratory is the imaging plate. For synchrotron sources, however, the read-out time is rate-limiting. This problem of detectors with insufficient data acquisition speeds is already substantial at third-generation sources; it can only get worse at fourth-generation ones. At several sites such as CHESS, the imaging plate is being succeeded by charge-coupled devices (CCDs) [6].

The second problem of data collection for macromolecules results from their inherent disorder. The size of the molecules means that they can no longer close pack in the crystal. As a result, the interstices are filled with disordered solvent, although water molecules close to the protein surface do take up an ordered structure. A typical protein crystal is ~50% 'solid' protein phase and ~50% liquid solvent [7]. The lattice forces are weak and often the surface residues of the protein show substantial static or thermal disorder. This exacerbates the weakness of the data at high resolution.

A knock-on effect of the high solvent content is the sensitivity of protein crystals to radiation damage. Although all crystals suffer to a small extent from the direct or primary damage, crystals with high solvent content experience a secondary effect caused by diffusion of the resulting radicals and ions through the solvent channels. This sensitivity of aqueous systems to X-rays is similar to the damage induced in living cells.

In this respect cryogenic freezing represents a fantastic advance. Cryogenic cooling has been applied for several years to small molecules, where the problem of freezing the sample is minimal and typically results in a decrease in average atomic temperature factor and in disorder. For macromolecules these advantages are also to be expected. However even more significant gains can be made in effective elimination of secondary radiation damage. Many protein crystals become in practice immortal in the X-ray beam, within the constraints described by Richard Henderson [8]. Of course, freezing of protein crystals [9] does bring experimental difficulties, involving the formation of ice either within the intermolecular interstices in the crystal or within any residual mother liquor around it. However, the problems are greatly outweighed by the gains. The great majority of crystals can now be frozen successfully after transfer to a cryoprotectant solution. The crystal is rapidly (seconds) scooped up in a fibre loop and exposed to the cold gas stream. For many proteins the process of freezing, while sometimes increasing the mosaic spread, reveals a resolution which could only be presumed to have been lost during the first exposure at room temperature. Experiments suggest that freezing can often increase the resolution by 0.5 Å or more.

Computing hardware, software and algorithms

Ever since the 1950s development in crystallography has relied upon, and indeed has driven, advances in computing. Structure solution and refinement has consistently required the most powerful computers available. The importance of computer power for atomic analysis of proteins lies in three areas of experimentation:

The first of these is data capture and reduction. The fast rates of data acquisition made possible by 2D detectors require substantial computing power and efficient algorithms for data reduction. The former has been effectively addressed through a number of computer programs. The latter has advanced through the development of several software packages, e.g. MOSFLM [10], DENZO [11] and XDS [12]. High-resolution data can be rapidly extracted from diffraction images with ease.

The second area is structure solution (the phase problem). This is particularly computer intensive, requiring ever more computer time for advanced algorithms in methods akin to maximum likelihood, pioneered by Gerard Bricogne [13,14], or direct methods such as those advanced by the groups of Woolfson [15], Hauptmann [16•] and Sheldrick [17]. For all of these methods the exponential increase in computer power has been vital, although it is still often limiting.

The final area is structure refinement. Algorithms for refinement of small molecules nowadays require minimal computing power; however, the needs increase rapidly as the size of the structure increases. Conventionally, macromolecular structures have been refined at resolutions well below atomic and have required the use of constrained [18] or restrained [19] procedures to either reduce the number of parameters refined or increase the observations using stereochemical restraints. The problems of limited resolution are highlighted in Table 2. With atomic-resolution data, it is possible to return to classic crystallographic approaches.

Procedures used in our refinements

For the structures that we are currently refining, a lower-resolution model has generally been available from multiple isomorphous replacement or molecular replacement at lower resolution. We have found the following protocol useful.

Starting model

The starting model is first refined against the atomic-resolution data using stereochemically restrained least-squares minimization, with conjugate gradient approximation, which gives a short computing time per cycle. We have used the PROLSQ suite [19] incorporating fast Fourier syntheses [20] from the CCP4 suite [21]. This is a subjective choice and it is clear that other programs such as TNT [22] or X-PLOR [23] should perform this role equally well.

The model is refined with isotropic atomic temperature factors. A serious problem arises when building water structure in proteins. It becomes increasingly time consuming to inspect subjectively the difference map as more and more 'waters' emerge in the improving density at atomic resolution. It is useful to apply at least a semi-objective set of criteria for water selection and we have used the Automatic Refinement Procedure (ARP) of Lamzin and Wilson [24,25], according to which water sites are selected on the basis of distance criteria, electron density and sphericity. This is not the only approach possible, but it is important that some such objective procedure is used. The B values of the waters selected are generally less than 50: all waters are accorded unit occupancy. ARP is used in combination with PROLSQ, but in the subsequent steps with SHELXL-93. Hydrogen atoms riding on their parent atoms are introduced into the model: this generally reduces the R factor by 0.5–1.0%.

The models refined using this procedure typically have R factors in the range 14–18%, which are comparable to those obtained for many proteins at resolutions of ~1.5 Å. One of the questions that we have sought to answer in our studies at atomic resolution was the source of this residual: for small structures, values in the range 2–4% are normal.

Anisotropic model

After convergence with PROLSQ, we next use the program SHELXL-93 [26••]. This incorporates many features that make it applicable to macromolecular structures. Particularly important is the existence of

Table 2. The effects of refining structures at resolution less than atomic.

Resolution (Å)	Protocol	Features identifiable
1.0	Full-matrix, anisotropic atoms	Fully resolved atoms
1.5	The border between isotropic/anisotropic	Hydrogens, disorder; ordered atoms distinguished
2.0	Isotropic atoms, stereochemical restraints	Some disorder
2.5	Isotropic model starts to break down	Shape of small groups
3.0	Rigid groups, some constraints	Shape of fragments, e.g. helices
6.0	Complete domains as rigid bodies	Globular protein

flexible options to model atomic anisotropy and of different kinds of restraints on geometry and on atomic temperature factors. Stereochemistry can be restrained against an external library based on small-molecule structures [27–29], or chemically identical units can be restrained to be the same. SHELXL-93 allows double conformations to be refined with complementary occupancies. The stereochemical restraints are only required for the disordered, highly flexible, regions: for well-ordered regions the X-ray terms are sufficient to define the atomic positions to within ~0.03 Å.

We usually run several cycles with isotropic temperature factors upon introducing the model to SHELXL-93. This rapidly converges to an R factor equivalent to that obtained with PROLSQ. Riding H atoms are included throughout. Subsequently, switching on refinement of anisotropic B values increases considerably the time per cycle. The conjugate gradient algorithm is used: this reduces the time per cycle but more cycles are required to achieve convergence. The R factor generally drops by 5–6% upon introducing anisotropy, leading to final values in the range 8–12%. Selection of waters can be made automatic, incorporating the anti-bumping restraints built into the program. We have continued to use ARP to achieve this.

Finally a few (2–5) cycles of full- or blocked-matrix minimization should be carried out. This is time consuming even for small proteins, and at present necessitates access to a supercomputer for large proteins. The time per cycle for the ribonuclease described below was more than 24 h·cpu on an SGI Challenge.

Validation of the refinement

To confirm the validity of introducing extra parameters for macromolecular structures, the most widely used criterion is the cross validation Rfree [30]. A random sample of at least 1000–2000 of the reflections is excluded from the least-squares and the residual for this provides Rfree. The introduction of extra parameters should produce a drop in Rfree as well as in R: at what point the drop becomes significant in absolute terms is less clear [31••].

At a resolution of, say, 1 Å , introducing anisotropic temperature factors is clearly valid. Rfree falls by almost as much as the R factor, about 5–6%. At 1.5 Å, Rfree will drop by 2–3%, provided the data are good, i.e. complete and with an Rmerge value of <25% in the outer shell: this is still significant. At 2 Å resolution, however, Rfree does not drop at all upon introduction of anisotropy (GM Sheldrick, personal communication) and the model should certainly be isotropic. Having established this protocol it seems unnecessary to assess anisotropy for all subsequent refinements at atomic resolution with Rfree.

Rfree has proved to be a much less useful validation tool in assessing the fine details of refinement [31••]. Such details include the modelling of alternative protein conformations and alternative models for water, for example, with or without solvent continuum, full or partial occupancies, or with the SHELXL-93 anti-bumping restraints as opposed to ARP. The difference in the real R factor, let alone Rfree, is generally minimal between the different models and does not provide an objective overall criterion. It is necessary to return to local criteria, such as the real-space R factor [32] and to detailed inspection of density to validate the models. Overfitting of the data must be avoided.

During the last cycles of refinement all data must be used, including those previously omitted from the computation of Rfree.

Non-crystallographic symmetry

The ratio of X-ray observables to parameters is sufficient that it is no longer necessary to impose non-crystallographic symmetry (NCS) restraints. Indeed, doing so might obscure small but significant differences between molecules in different environments in the cell, which will provide general guidelines about how tightly NCS restraints should be imposed at lower resolution.

Results

Structures with atomic-resolution data

At the EMBL in Hamburg data to atomic resolution have been collected on several proteins, although almost all of the results remain to be published. A representative list derived from our own studies and those of visitors to the EMBL is given in Table 3. These studies were initiated after installation of the first imaging-plate scanner. The first structure to be studied as a test for the system was a medium-sized molecule, β-cyclodextrin. It was refined to ~3% R factor at 0.9 Å resolution, i.e. much better resolution than obtained in previous work (previous studies of cyclodextrins had resulted in data to lower resolution, many with R factors in the range 7–15%), setting a benchmark for the approach.

Advantages of atomic-resolution data in refinement

Atomic-resolution refinement can be expected to afford the researcher many advantages:

(a) The final R factor should lie within the range of 8–12% for the anisotropic model including riding H atoms. The Rfree is usually 2–4% higher. Moreover, the increase in R factor with resolution is substantially less than for the isotropic model. This has clear implications for the accuracy of the atomic positions [31••,33,34] (DWJ Cruickshank, personal communication).

(b) The difference Fourier synthesis should have a root mean square of 0.05 e $Å^{-3}$, with the largest features not exceeding ±0.4 e $Å^{-3}$. The 'cleanness' of the maps allows much easier visualization of solvent, disordered residues and H atoms.

(c) The positions of C, N and O atoms in the ordered parts of the structure will have estimated errors

Table 3. Representative proteins for which data have been recorded at atomic resolution at the EMBL.

Protein	Origin of research group	Asymmetric unit (kDa)	Vm ($Å^3Da^{-1}$)	Resolution (Å)	Current R factor
Rubredoxin Dv	Seattle	6.0	1.6	1.00 (room)	8.7
Rubredoxin Dv	Seattle	6.0	1.6	0.92 (room)	7.4
Rubredoxin Cp Fe	Grenoble	6.0	2.2	1.10 (room)	9.0
Rubredoxin Cp Zn	Grenoble	6.0	2.2	1.20 (room)	10.7
BPTI	Hamburg	6.5	1.9	1.08 (room)	10.5
ROP	Heraklion	7.0	1.7	1.08 (room)	10
Protein G	Leicester	6.6	2.2	1.10 (room)	9.7
Cytochrome c_6	Lisbon	9.3	2.3	1.10 (room)	NA
Insulin	Hamburg	11.0	1.8	1.20 (room)	13
Insulin	Hamburg	11.0	1.8	0.98 (100K)	13
RNase P1	Moscow	11.0	2.4	1.08 (room)	11.0
Lysozyme P1	Seattle	14.3	1.7	0.92 (100K)	17 (iso)
Cutinase	Marseille	NA	NA	1.05 (room)	9.5
Trypsin bacterial	NOVO	19.2	1.9	1.10 (room)	7.5
Trypsin bacterial	NOVO	19.2	1.9	1.02 (180K)	10.8
Trypsin bacterial	NOVO	19.2	1.9	0.96 (120K)	8.9
Trypsin bacterial	NOVO	19.2	1.9	0.98 (90K)	10.0
RNase Sa	Bratislava	21.1	2.3	1.20 (room)	10.6
RNase Sa recomb	Bratislava	21.1	2.3	1.00 (110K)	NA
RNase Sa GMP	Bratislava	21.1	2.3	1.15 (room)	10.9
Xylanase	NOVO	21.8	NA	1.00 (100K)	NA
Trypsin fungal	NOVO	22.1	2.0	1.07 (room)	11.0
Trypsin fungal	NOVO	22.1	2.0	0.93 (100K)	11.0
Savinase	NOVO	26.7	1.8	1.00 (110K)	10.7
Subtilisin BPN'	P&G	27.5	2.1	1.15 (100K)	NA
LADH Zn	Stockholm	79.6	2.4	1.10 (110K)	NA
LADH Cd	Stockholm	79.6	2.4	1.15 (110K)	NA
LADH Cd/DMSO	Stockholm	79.6	2.4	1.00 (110K)	NA

Full details of the research groups are not provided because none of the structures have yet been published. NOVO represents Novo-Nordisk of Copenhagen, P&G Proctor and Gamble. The temperature at which atomic-resolution data were obtained are indicated (room, room temperature).

of ~0.03 Å. The average for the whole structure may be 0.05 Å, reflecting the substantial errors in the disordered residues.

(d) The significance of the anisotropic atomic model is clear from the R factors. In addition, analysis of the correlation of thermal vibration with temperature and with distance from the centre of gravity of the molecule (TR Schneider, KS Wilson and F Parak, abstract MO70, Meeting of the American Crystallographic Association, Montreal, 1995) confirms that the B values are physically meaningful.

(e) The positions of many H atoms can be seen in the density maps: for rubredoxin, for example, about two thirds of them could be directly picked up using SHELXL-93.

(f) Dual conformations can be identified for a greater proportion of residues: ~10% in our experience. Serines, methionines, lysines, threonines and valines are especially prone to this. Dual conformations have even been observed for tryptophan, and for a tyrosine with two complementary H bond networks. Even at atomic resolution, however, a few residues continue to reveal no clearly identifiable average conformation.

(g) The water structure emerges from the continuum. As the residuals in the difference Fourier decrease, significant features emerge at increasing distances from the protein surface.

The refinement of the bacterial ribonuclease from *Streptomyces aureofaciens* illustrates in more detail most of the points made here [35].

Ab initio phasing of atomic-resolution data

This has so far only worked for proteins such as avian pancreatic polypeptide [14], crambin [15], rubredoxin [16•] and cytochrome c_6 [36••,37] which contain heavy-atom units such as metals or disulphide bridges. Cytochrome c_6 is the first protein with previously unknown structure that has succumbed to this approach. So far, no all-light atom macromolecular structures have been solved *ab initio*.

For most of these studies the heavy atoms were first located from the Patterson synthesis, and the initial phase set extended to reveal the bulk of the rest of the structure. For rubredoxin and avian pancreatic polypeptide, classical direct methods were sufficient to produce a meaningful, almost complete model.

George Sheldrick and Durward Cruickshank have both pointed out why the presence of heavy atoms confers such an advantage: the B values of the heavy atoms are so much less than the rest of the structure that they dominate the scattering factors at high resolution, where the lighter solvent and surface residues become insignificant. Thus the structure to be initially defined at atomic resolution becomes rather small and easier to produce via direct methods. It may be a little while before an all-light atom macromolecule is solved *ab initio*.

Future prospects

When we started this work, we did not have great expectations that any but a tiny number of protein crystals (rubredoxin and lysozyme!) would really provide atomic-resolution data. However, the combination of bright X-ray sources, efficient 2D detectors and cryogenic temperatures has produced a rapidly growing number of structures, and not all of them are small tightly packed systems such as rubredoxin. For example, in Hamburg we have recorded data to 1.1 Å resolution for liver alcohol dehydrogenase, which has two 80 kDa monomers in the asymmetric unit, Elspeth Garman from Oxford has data to 1.0 Å on neuraminidase (also from Hamburg) and Jeremy Tame from York has 1.1 Å data on an oligopeptide-binding protein with a molecular mass of 80 kDa measured at the SRS in Daresbury, UK.

The number of atomic-resolution structures solved will increase and within the next two years several coordinate sets will be deposited in the Macromolecular Structural Database (MSD; currently called the Protein Data Bank [38]) at Brookhaven, providing a wealth of structural information for analysis and comparison of the details of protein structure. Far from simplifying the process of protein structure prediction, such studies are revealing the rules of protein structure to be even more complex than previously envisaged. The proportion of residues for which multiple conformations can be seen is greater at atomic resolution, further complicating the lives of those attempting structure prediction.

The use of fast detectors such as CCDs will without doubt transform the field. For atomic-resolution data on bacterial trypsin, for example, the total exposure time is ~30 min, but data collection takes about 6 h because of the time required for read out of the image plate. The collection of such data in less than 1 h will allow studies to be extended to series of complexes which, together with the use of flash freezing, should allow detailed probing of the enzyme mechanism in tractable times, and at resolutions approaching those of real interest to chemists.

The information content of these data sets is phenomenal. It is clear that we cannot at the moment necessarily expect to extract all of the relevant data to produce a definitive model of, say, water structure. Improved algorithms will certainly be developed as more and more models become available. It is crucial that the raw data are not lost but are archived in the MSD. Journal editors must be persuaded that the X-ray data as well as the model should be deposited as a condition of publication.

Direct or Patterson solutions of small metalloproteins is now a reality. We can hope that *ab initio* solution of normal proteins via maximum likelihood will become a reality during the next years [13]. Restating one of our initial observations, the collection of data to atomic resolution makes all subsequent steps in a crystal structure analysis and refinement easier. It should be pursued whenever possible.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Glover I, Haneef I, Pitts J, Wood S, Moss D, Tickle I, Blundell T: **Conformational flexibility in a small globular hormone: X-ray analysis of avian pancreatic polypeptide at 0.98 Å resolution.** *Biopolymers* 1983, **22**:293–304.
2. Teeter MM, Roe SM, Heo NH: **Atomic resolution (0.83 Å) crystal structure of the hydrophobic protein crambin at 130 K.** *J Mol Biol* 1993, **230**:292–311.
3. Watenpaugh KD, Sieker LC, Jensen LH: **Crystallographic refinement of rubredoxin at 1.2 Å resolution.** *J Mol Biol* 1980, **138**:615–633.
4. Chang W, Stuart D, Dai J, Todd R, Zhang J, Xie D, Kuang B, Liang D: **Refinement of the crystal structure of insulin at 1.2 Å resolution.** *Sci Sin* 1986, **29**:1273–1284.
5. Sheldrick GM: **Phase annealing in SHELX-90: direct methods for larger structures.** *Acta Crystallogr [A]* 1990, **46**:467–473.
6. Gruner SM, Ealick SE: **Charge coupled device X-ray detectors for macromolecular crystallography.** *Structure* 1995, **3**:13–15.
7. Matthews BW: **Solvent content of protein crystals.** *Acta Crystallogr* 1968, **33**:491–497.
8. Henderson R: **Cryo protection of protein crystals against radiation damage in electron and X-ray diffraction.** *Proc R Soc Lond [B]* 1990, **241**:6–8.
9. Watenpaugh KD: **Macromolecular crystallography at cryogenic temperatures.** *Curr Opin Struct Biol* 1991, **1**:1012–1015.
10. Leslie AGW: **Recent changes to the MOSFLM package for processing film and image plate data.** *CCP4 and ESF-EACBM Newsletter on Protein Crystallography.* Daresbury: SERC Laboratory; 1993.
11. Otwinowski Z, Minor W: *DENZO: A Film Processing Program for Macromolecular Crystallography.* New Haven: Yale University; 1993.
12. Kabsch W: **Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector.** *J Appl Crystallogr* 1988, **21**:916–924.
13. Bricogne G: **Direct phase determination by entropy maximization and likelihood ranking: status report and perspectives.** *Acta Crystallogr [D]* 1993, **49**:37–60.
14. Carter CW: **Entropy, likelihood and phase determination.** *Structure* 1995, **3**:147–150.

15. Woolfson MM, Yao J-X: **On the application of phase relationships to complex structures. XXX. *Ab initio* solution of a small protein by SAYTAN.** *Acta Crystallogr [A]* 1990, **46**:409–413.

16. • Weeks CM, Hauptman HA, Smith GD, Blessing RH, Teeter MM, Miller R: **Crambin: a direct solution for a 400-Atom structure.** *Acta Crystallogr [D]* 1995, **51**:33–38.
The application of the 'shake-n-bake' procedure to the *ab initio* phasing of a small protein, that lacks a metal atom but possesses only sulphur atoms in disulphide bridges.

17. Sheldrick GM, Dauter Z, Wilson KS, Hope H, Sieker LC: **The application of direct methods and patterson interpretation to high-resolution native protein data.** *Acta Crystallogr [D]* 1993, **49**: 18–23.

18. Sussman JL, Holbrook SR, Church GM, Kim S-H: **A Structure-factor least-squares refinement procedure for macromolecular structures using constrained and restrained parameters.** *Acta Crystallogr [A]* 1977, **33**:800–804.

19. Konnert JH, Hendrickson WA: **A Restrained-parameter thermal-factor refinement procedure.** *Acta Crystallogr [A]* 1980, **36**:344–350.

20. Agarwal RC: **A new least-squares refinement technique based on the fast Fourier Transform algorithm.** *Acta Crystallogr [A]* 1978, **34**:791–809.

21. CCP4: **Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography.** *Acta Crystallogr [D]* 1994, **50**:760–763.

22. Tronrud DE, Ten Eyck LF, Matthews BW: **An efficient general-purpose least-squares refinement program for macromolecular structures.** *Acta Crystallogr [A]* 1987, **43**:489–501.

23. Brünger AT, Kuriyan J, Karplus M: **Crystallographic R factor refinement by molecular dynamics.** *Science* 1987, **235**:458–460.

24. Lamzin VS, Wilson KS: **Automated refinement of protein models.** *Acta Crystallogr [D]* 1993, **49**:129–147.

25. Lamzin VS, Wilson KS: **Automated refinement for protein crystallography.** *Methods Enzymol* 1996, in press.

26. •• Sheldrick GM, Schneider T: **SHELXL: High-resolution refinement.** *Methods Enzymol* 1996, in press.
The paper describes how to use most effectively the SHELXL programme to refine atomic-resolution models of macromolecules. It is essential reading for those embarking on such studies.

27. Engh RA, Huber R: **Accurate bond and angle parameters for X-ray protein structure refinement.** *Acta Crystallogr [A]* 1991 **47**:392–400.

28. Priestle JP: **Stereochemical dictionaries for protein structure refinement and model building.** *Structure* 1994, **15**:911–913.

29. Lamzin VS, Dauter Z, Wilson KS: **Dictionary of protein stereochemistry.** *J Appl Crystallogr* 1995, **28**:338–340.

30. Brünger AT: **Assessment of phase accuracy by cross validation: the free R value. Methods and application.** *Acta Crystallogr [D]* 1993, **49**:24–36.

31. •• Dodson EJ, Kleywegt K, Wilson KS: **Report of a workshop on the use of statistical validators in protein X-ray crystallography.** *Acta Crystallogr [D]* 1996, in press.
This workshop discussed the need for statistical validators of atomic models with regard to the experimental X-ray data. Several procedures were recommended and a number of pitfalls identified.

32. Brändén C-I, Jones TA: **Between objectivity and subjectivity.** *Nature* 1990, **343**:687–689.

33. Luzzati V: **Traitement statistique des erreurs dans la determination des structures cristallines.** *Acta Crystallogr* 1952, **5**:802–819.

34. Read RJ: **Improved Fourier coefficients for maps using phases from partial structures with errors.** *Acta Crystallogr [A]* 1986, **42**:140–149.

35. Sevcik J, Dauter Z, Lamzin VS, Wilson KS: **Ribonuclease from *Streptomyces aureofaciens* at atomic resolution.** *Acta Crystallogr [D]* 1995, **51**: in press.

36. •• Frazao C, Soares CM, Carrondo MA, Pohl E, Dauter Z, Wilson KS, Hervás M, Navarro JA, De la Rosa MA, Sheldrick GM: ***Ab initio* determination of the crystal structure of cytochrome c6; comparison with plastocyanin.** *Structure* 1995, in press.
The first *ab initio* solution of a previously unknown protein structure obtained from amplitudes alone. Again, the importance of heavier atoms, an iron and three sulphurs, is emphasized.

37. Sheldrick GM, Gould RO: **Structure solution by iterative peaklist optimization and tangent expansion in space group P1.** *Acta Crystallogr [D]* 1995, in press.

38. Bernstein FC, Koetzle TF, Williams GJB, Mayer EF, Bryce MD, Rodgers JR, Kennard O, Simanouchi T, Tasumi M: **The Protein Data Bank: a computer-based archival file for macromolecular structures.** *J Mol Biol* 1977, **112**:535–542.

Z Dauter and KS Wilson, European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany and Department of Chemistry, University of York, Heslington, York YO1 5DD, UK.
Z Dauter e-mail: dauter@yorvic.york.ac.uk
KS Wilson e-mail: keith@yorvic.york.ac.uk
VS Lamzin, European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, D-22603 Hamburg, Germany.
e-mail: victor@embl-hamburg.de